Attacks on hash functions: Cat 5 storm or a drizzle?

Ilya Mironov
Microsoft Research, Silicon Valley Campus

September 15, 2005

Outline

Hash functions:

- Definitions
- Constructions
- Attacks
- What to do

Outline

Hash functions:

- Definitions
- Constructions
- Attacks
- What to do

Functions of Hash Functions

- Entropy extractors
- Randomization
- Cryptography
  - Signatures
  - Authentication

Entropy extraction

Alice $g \in G$, ord $g = p$ Bob

random a

g^a →

random b

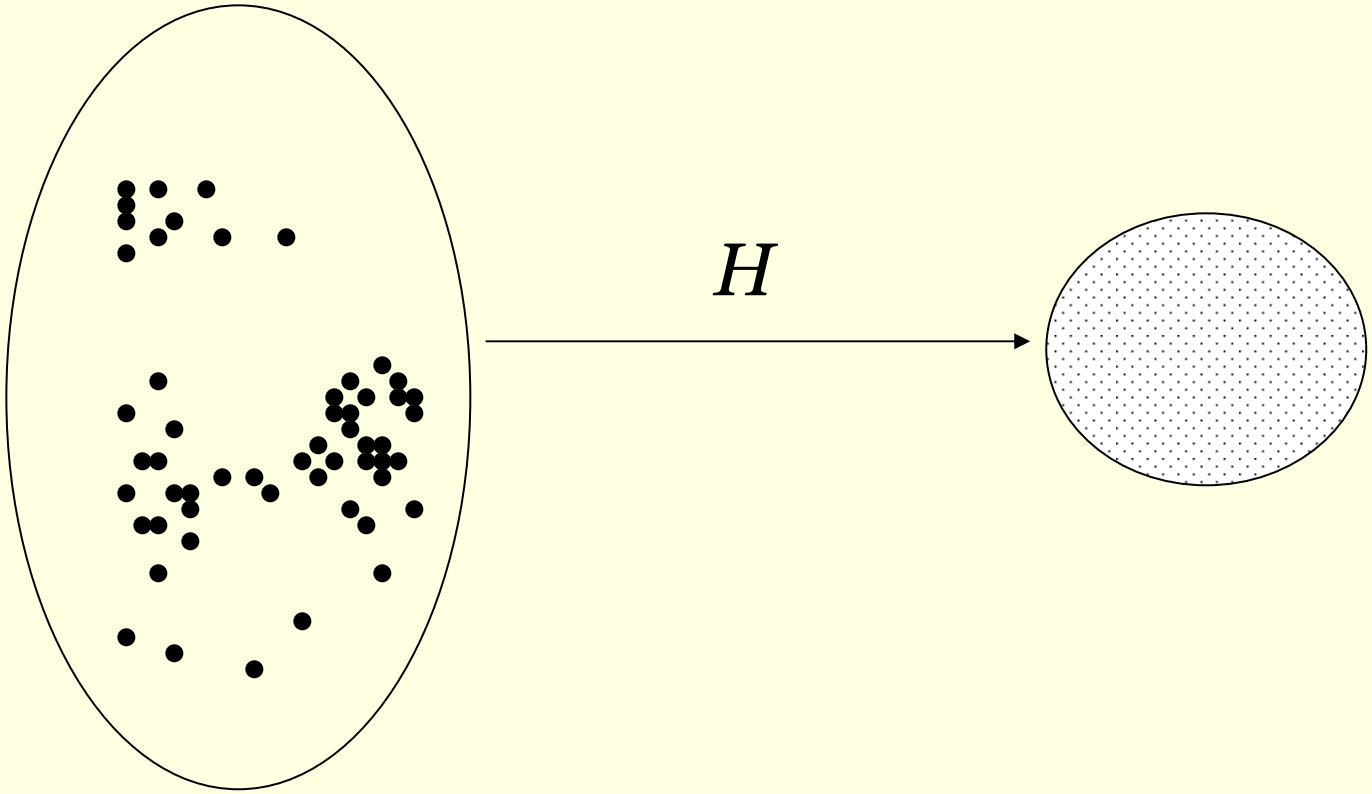
← g^b

compute g^{ab} compute g^{ab}

shared key $= H(g^{ab})$

Enrtopy extraction

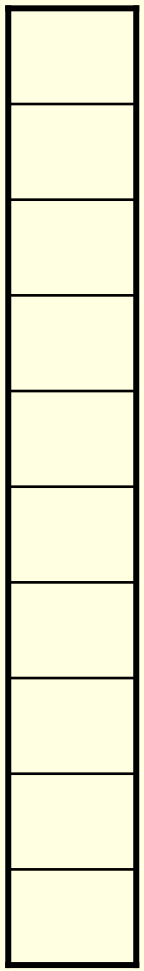
IP, time, name, MAC address $\rightarrow$ GUID



Randomization

- Scenario: hash table

3 12 78 40 67 15



Cryptography

RSA signature:

parameters: $N = pq$, $ed = 1 \bmod \varphi(N)$

sign $C = H(M)^d \bmod N$, **verify** $C^e = H(M) \bmod N$

Attack:

sign M_1, M_2, obtain C_1, C_2

$(C_1C_2)^d = H(M_1)H(M_2)$

Authentication

Alice

Bob

shared K

M

$H_K(M)$

What is a hash function?

$$H: \{0,1\}^* \rightarrow \{0,1\}^n$$

What is a hash function?

$$H_k: \{0,1\}^* \rightarrow \{0,1\}^n$$

What is cryptographically strong hash function?

- Collision resistant:

  Find x,y, so that $H(x)=H(y)$

- Second preimage-resistant:

  Given: x

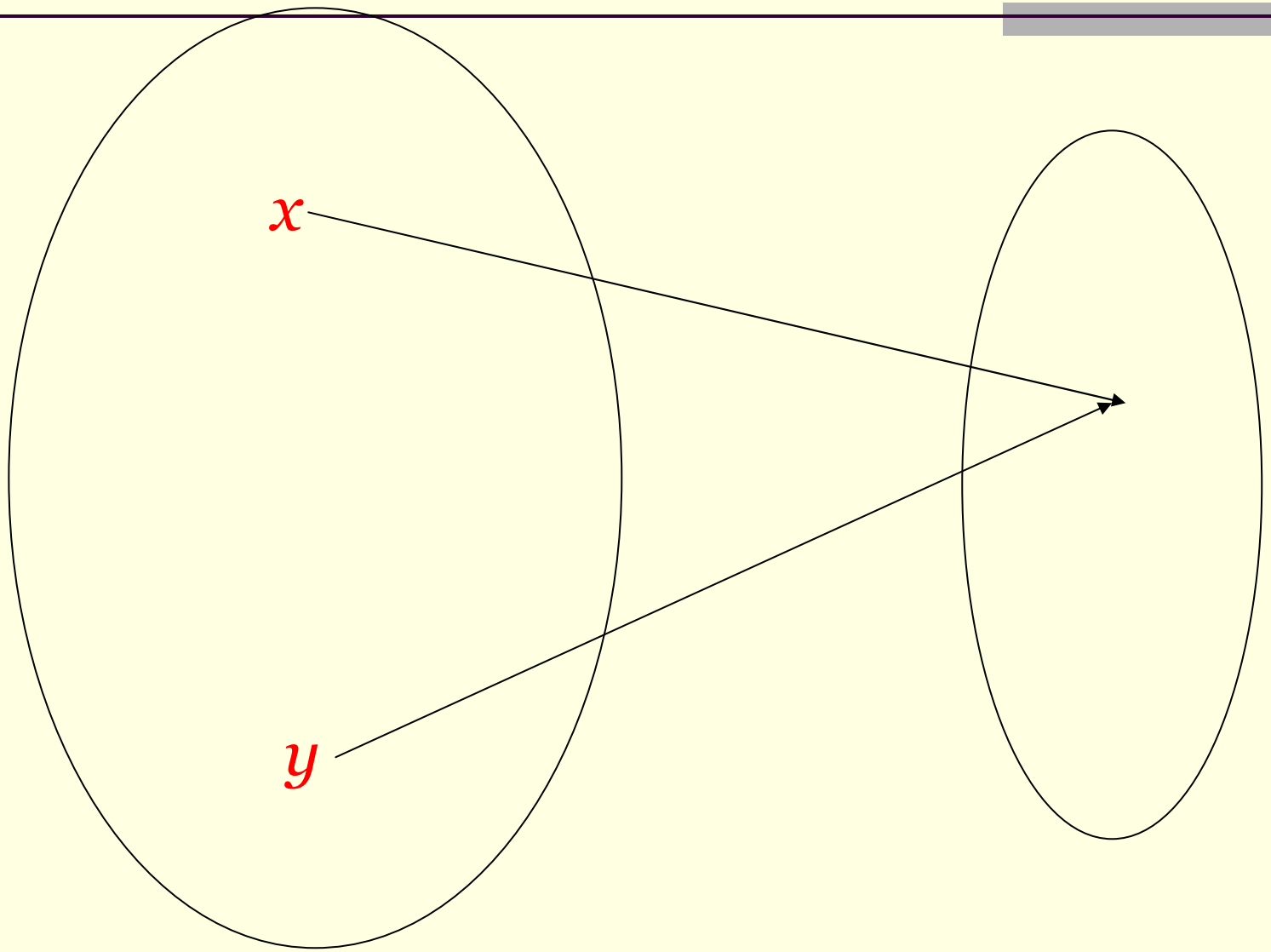
  Find: y, so that $H(x)=H(y)$

- One-way:

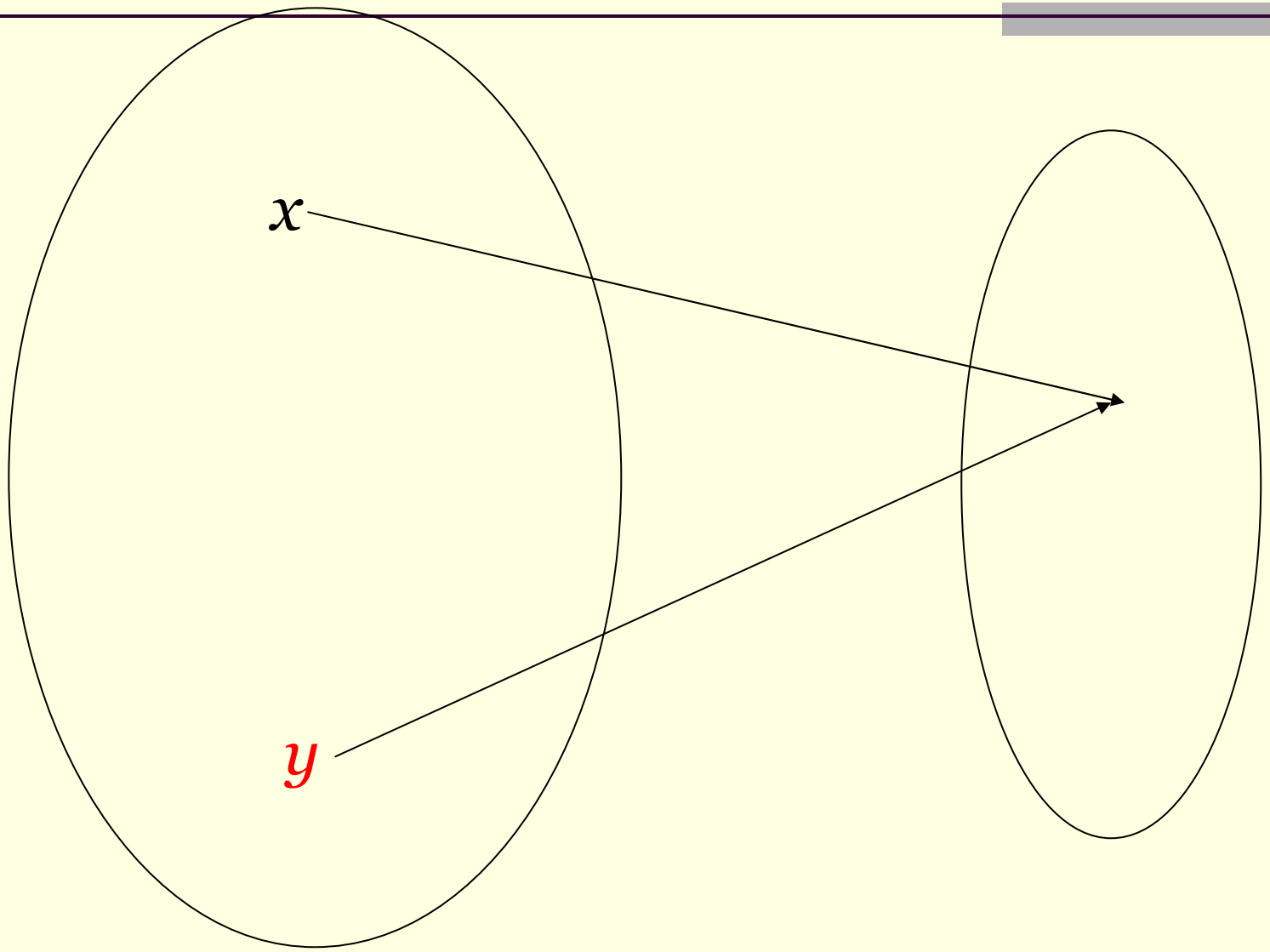
  Given: $z=H(x)$

  Find: y, so that $H(y)=z$

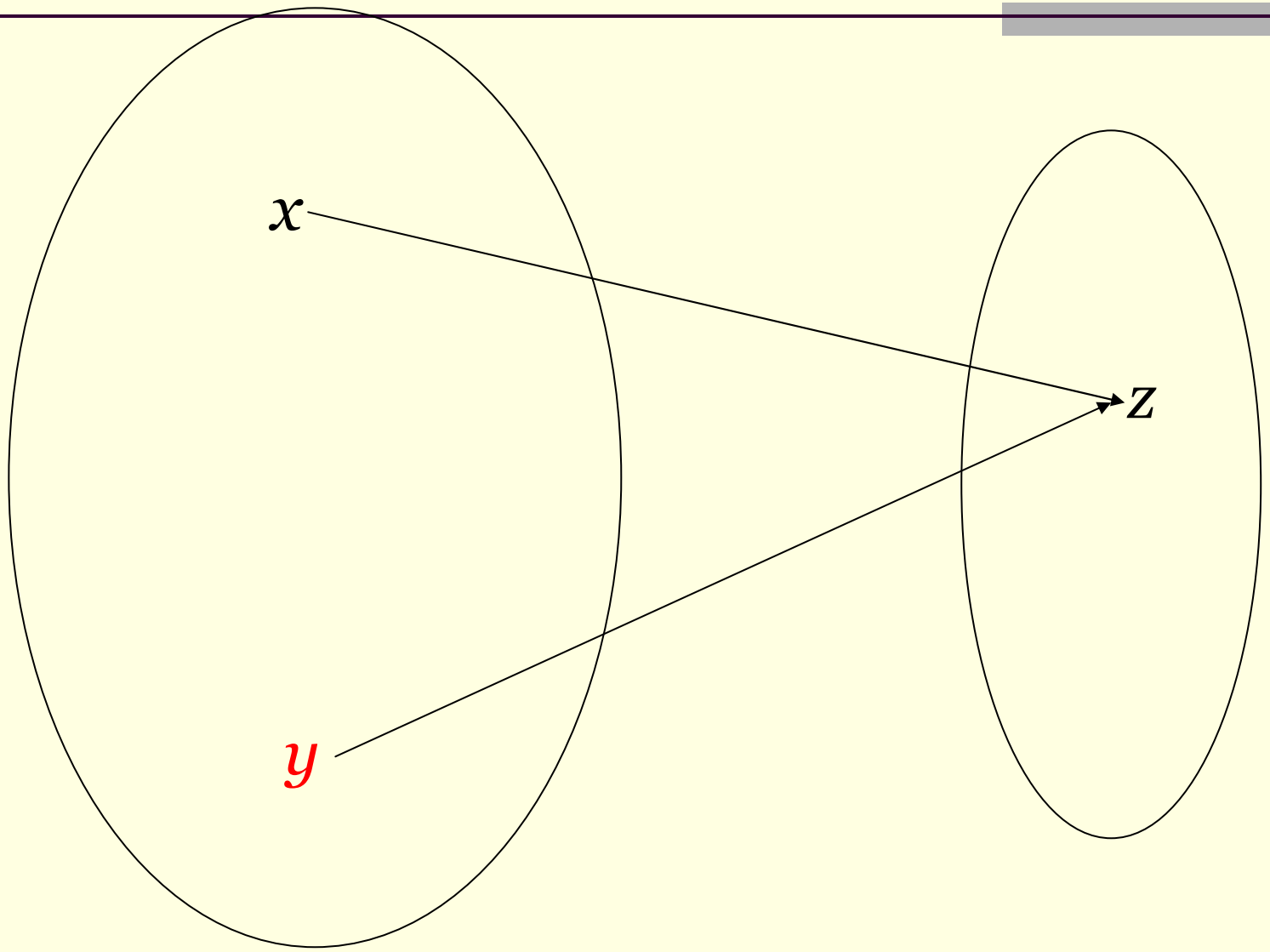
Collision resistant



Second preimage-resistant



One-way

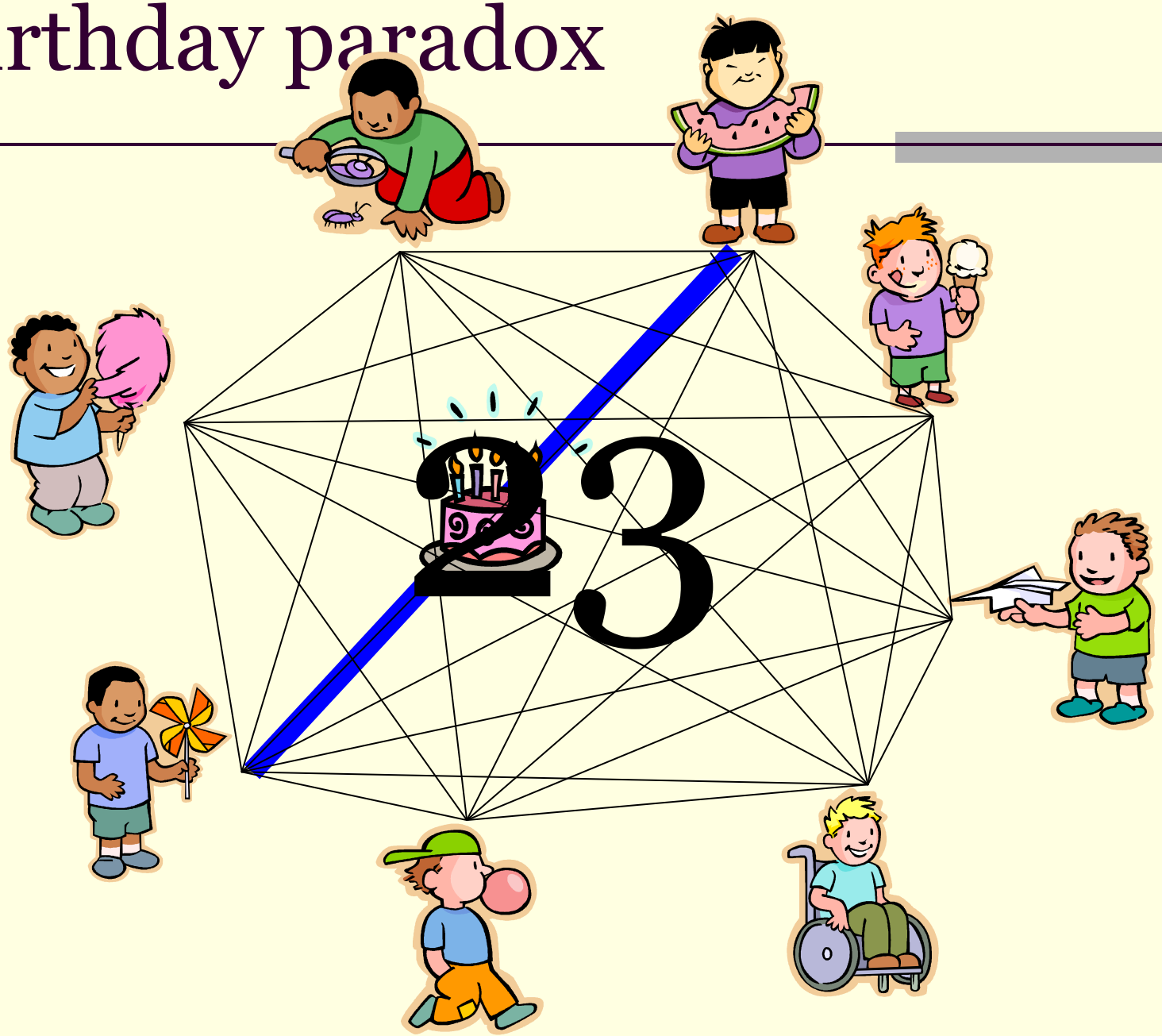


Attacks

- Collision-resistance:
  - Compute M_1, M_2, so that $H(M_1) = H(M_2)$
  - Sign M_1
  - M_2 is signed as well!

- Second preimage-resistance:
  - Find signature on M_1
  - Compute M_2, so that $H(M_1) = H(M_2)$
  - M_2 is signed too!

Birthday paradox



Theoretical boundaries

- Collisions
  - Always exist
  - Birthday paradox: $2^{n/2}$
- Second preimage 2^n
- One-wayness 2^n

Dan Simon'98

Outline

Hash functions:

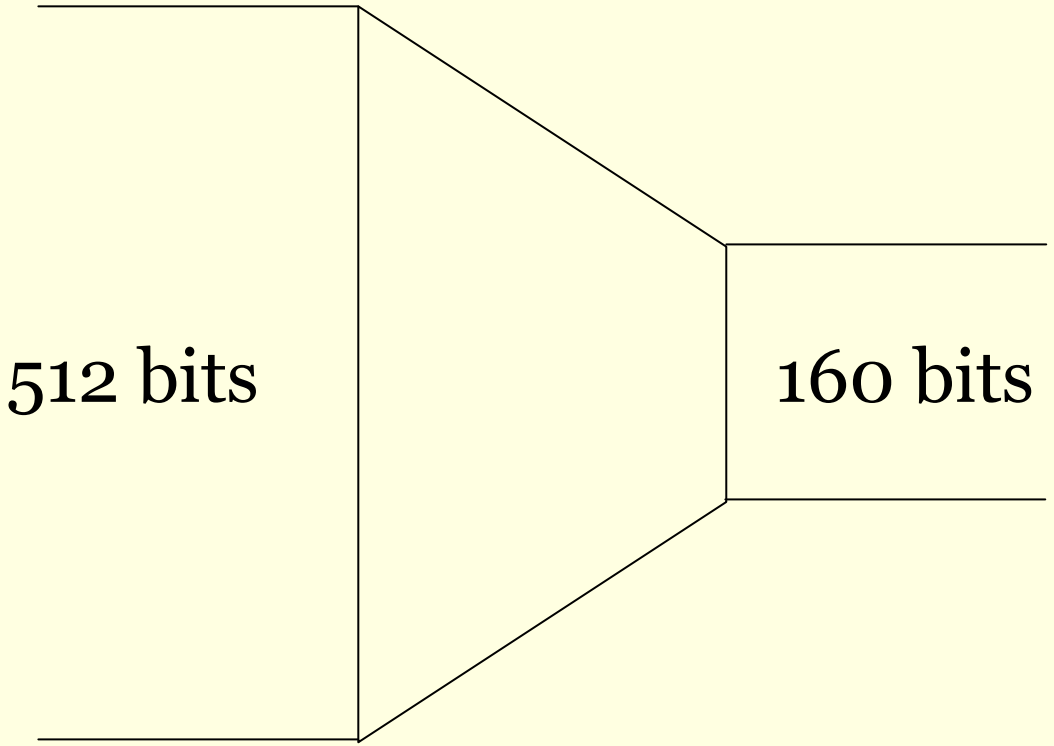
- Definitions
- Constructions
- Attacks
- What to do

Constructions

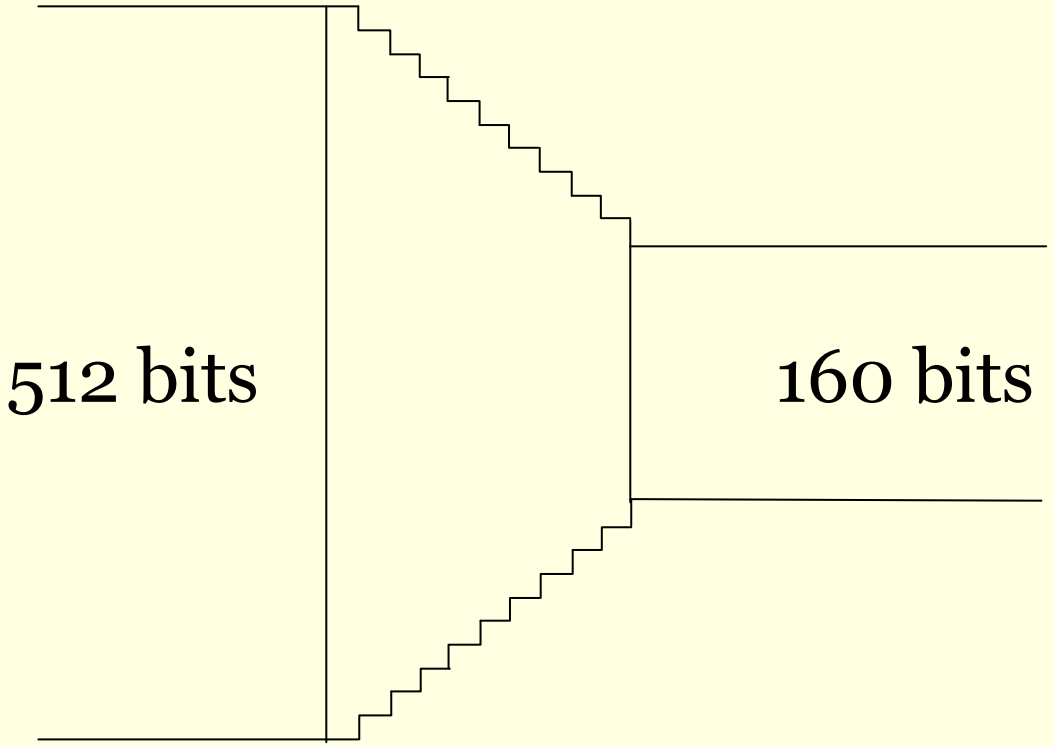
Two main ingredients

- Fixed compression function
- Domain extender

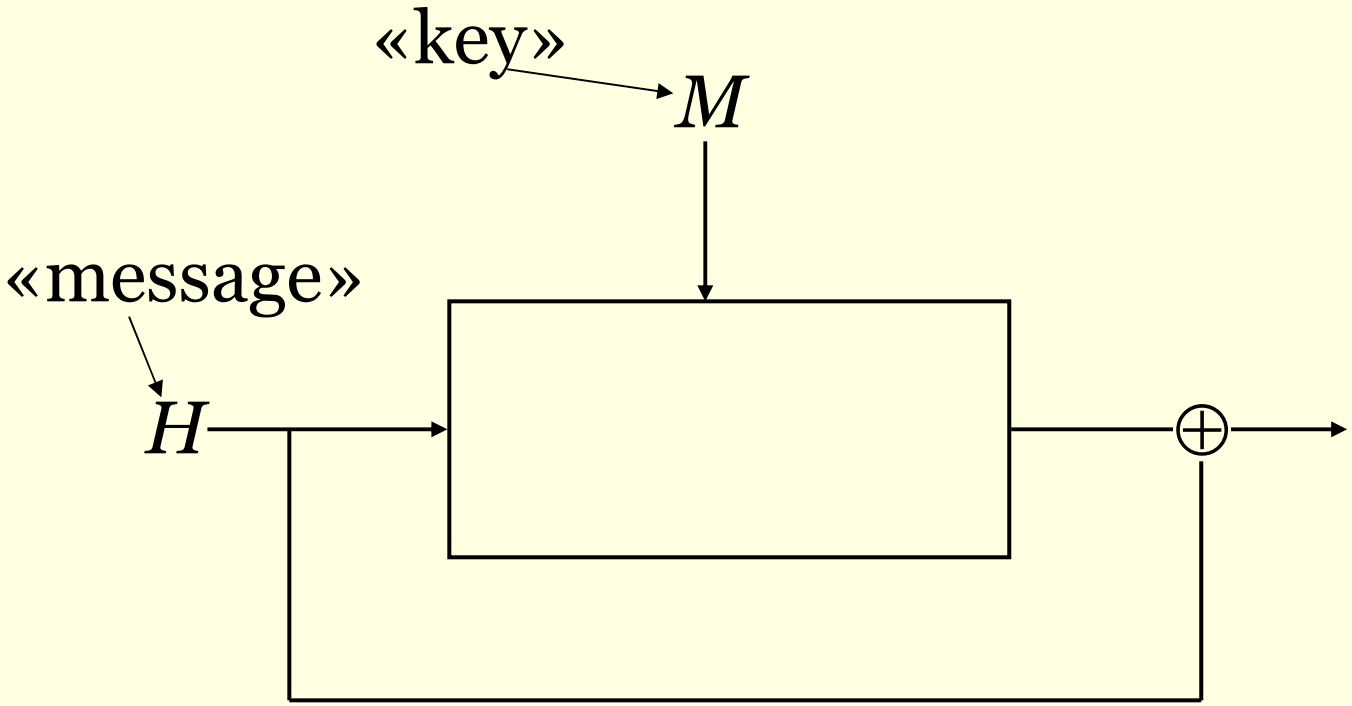
Compression function



Compression function

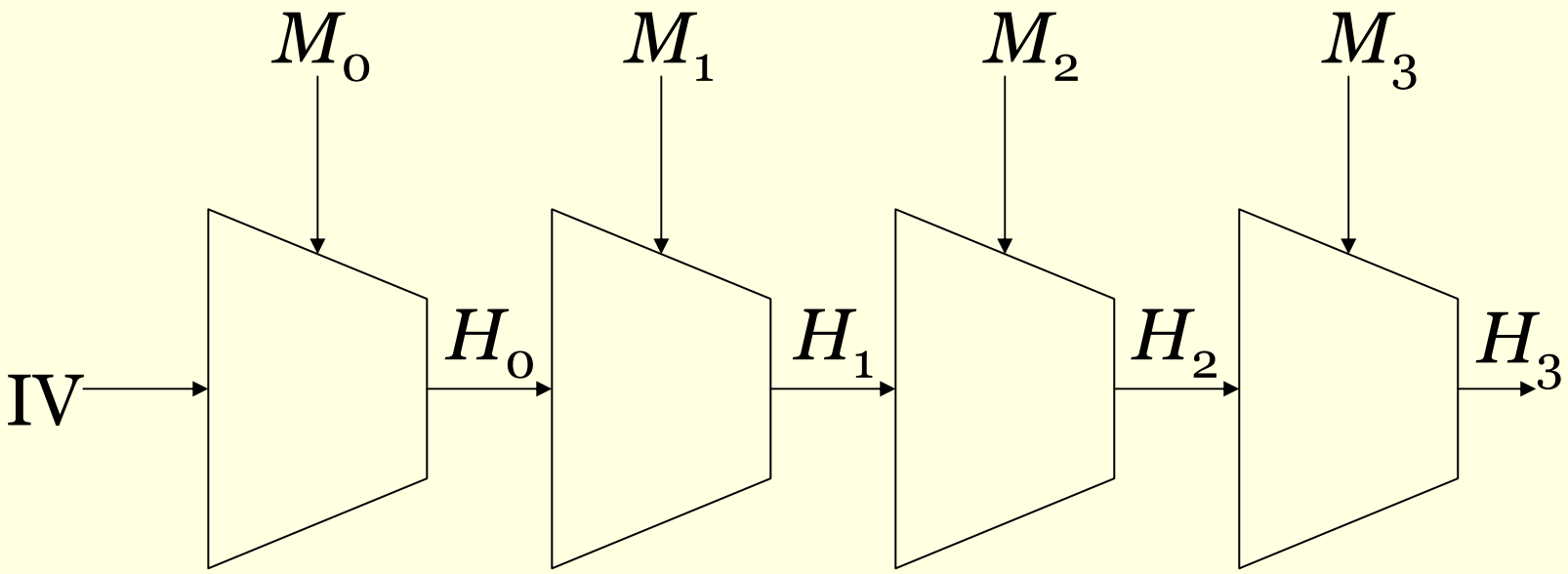


Compression function



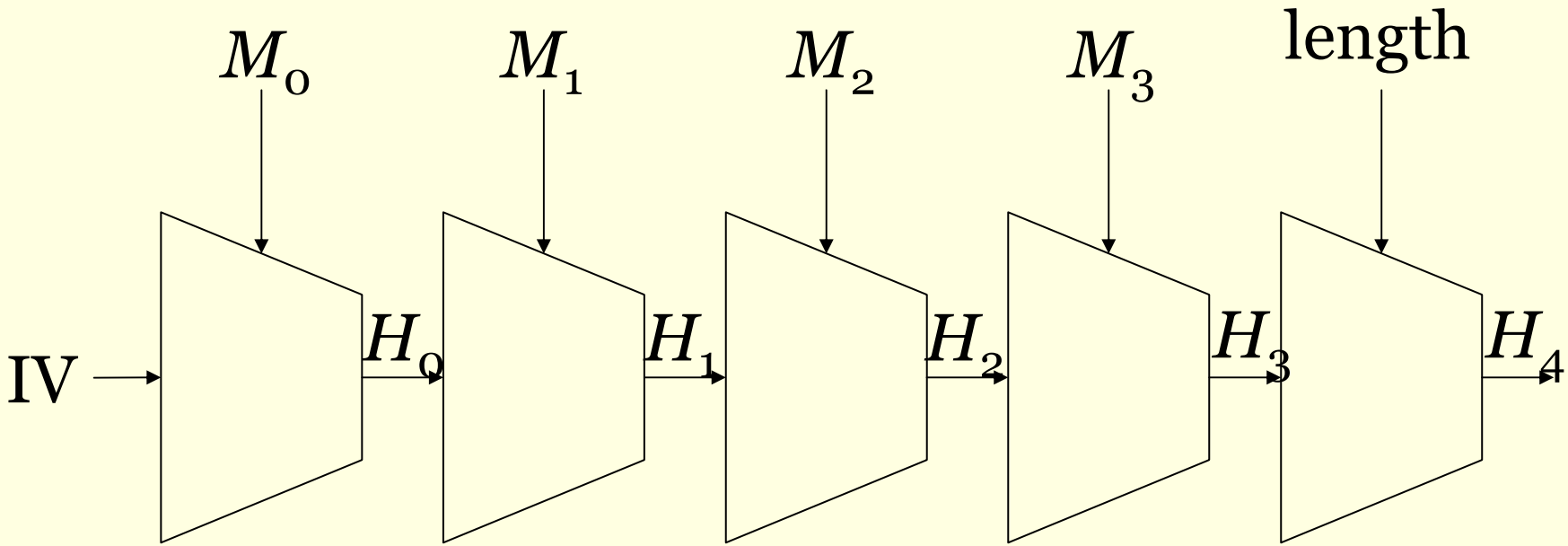
Davies-Meyer construction (~1979)

Domain extender

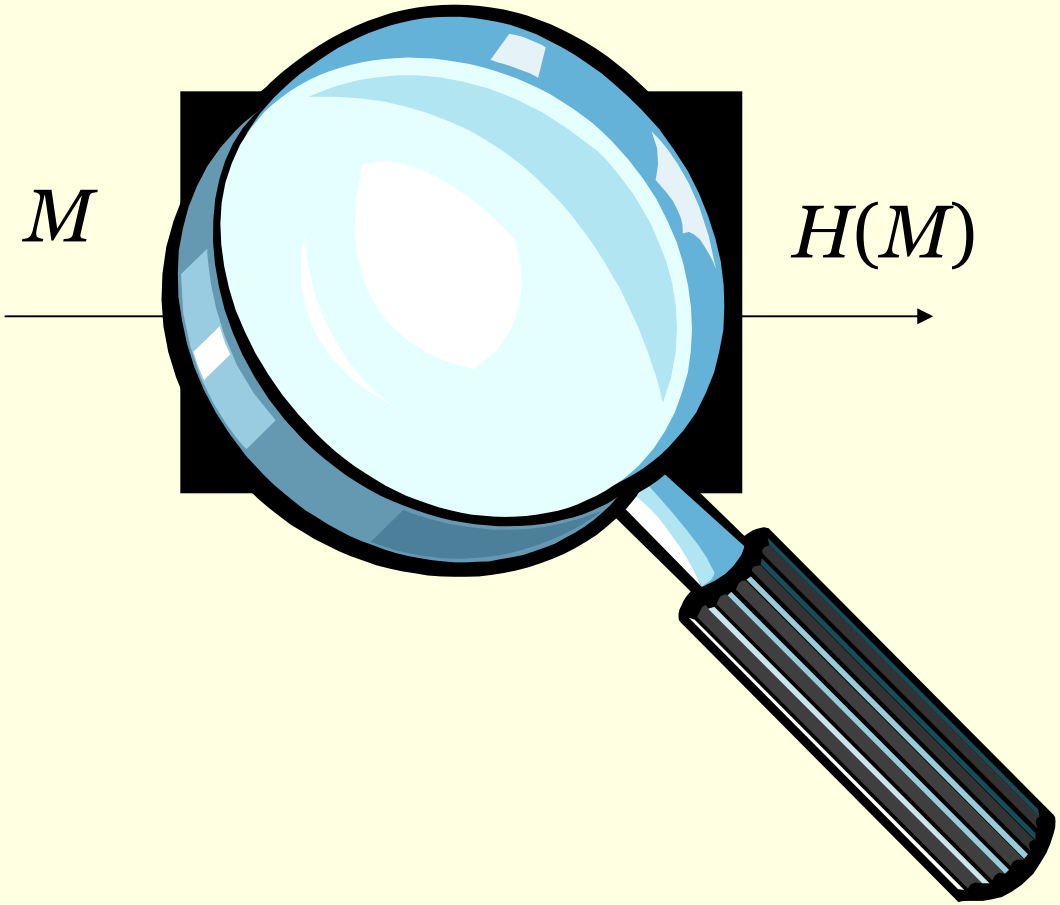


Merkle-Damgård construction (1989)

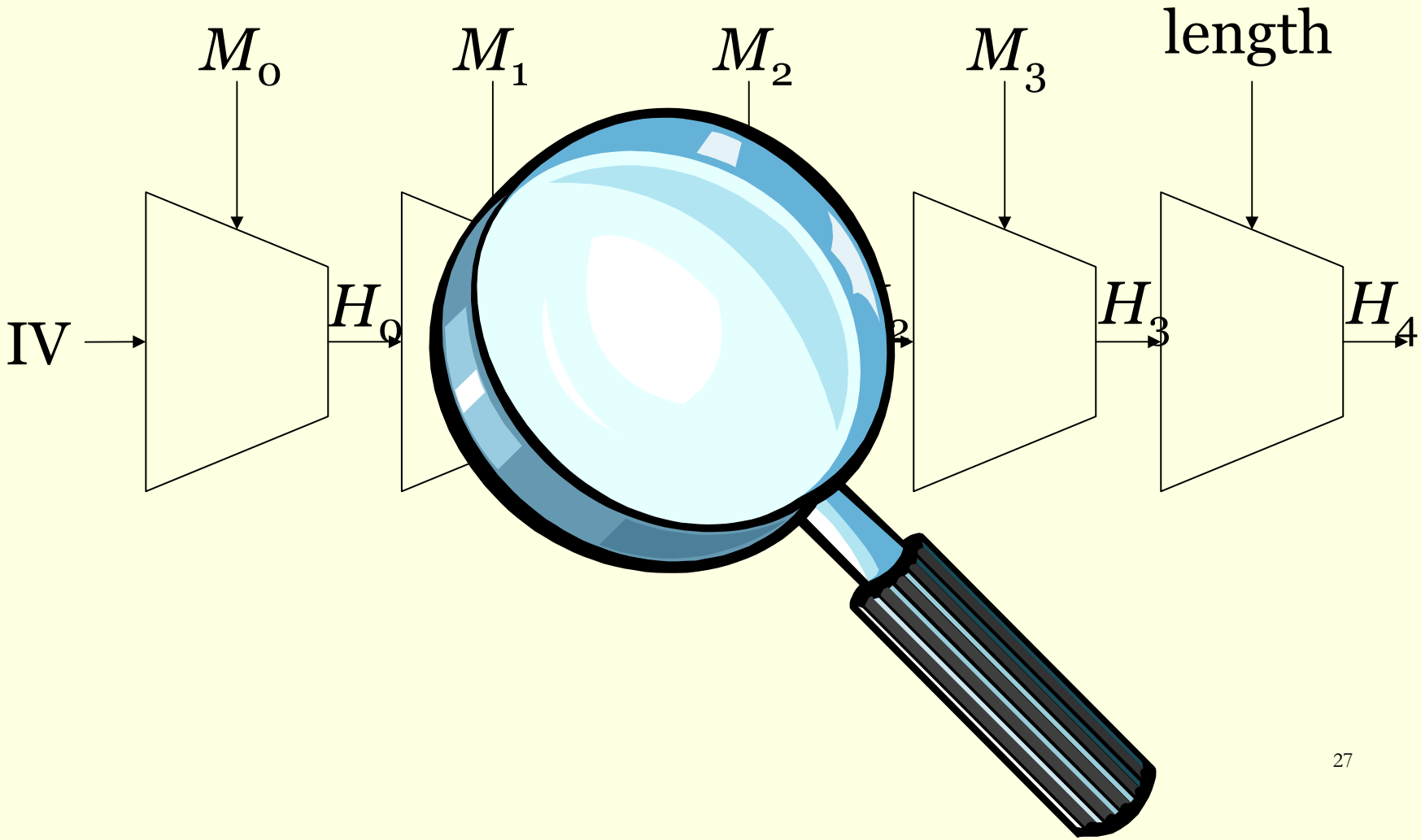
Proof?



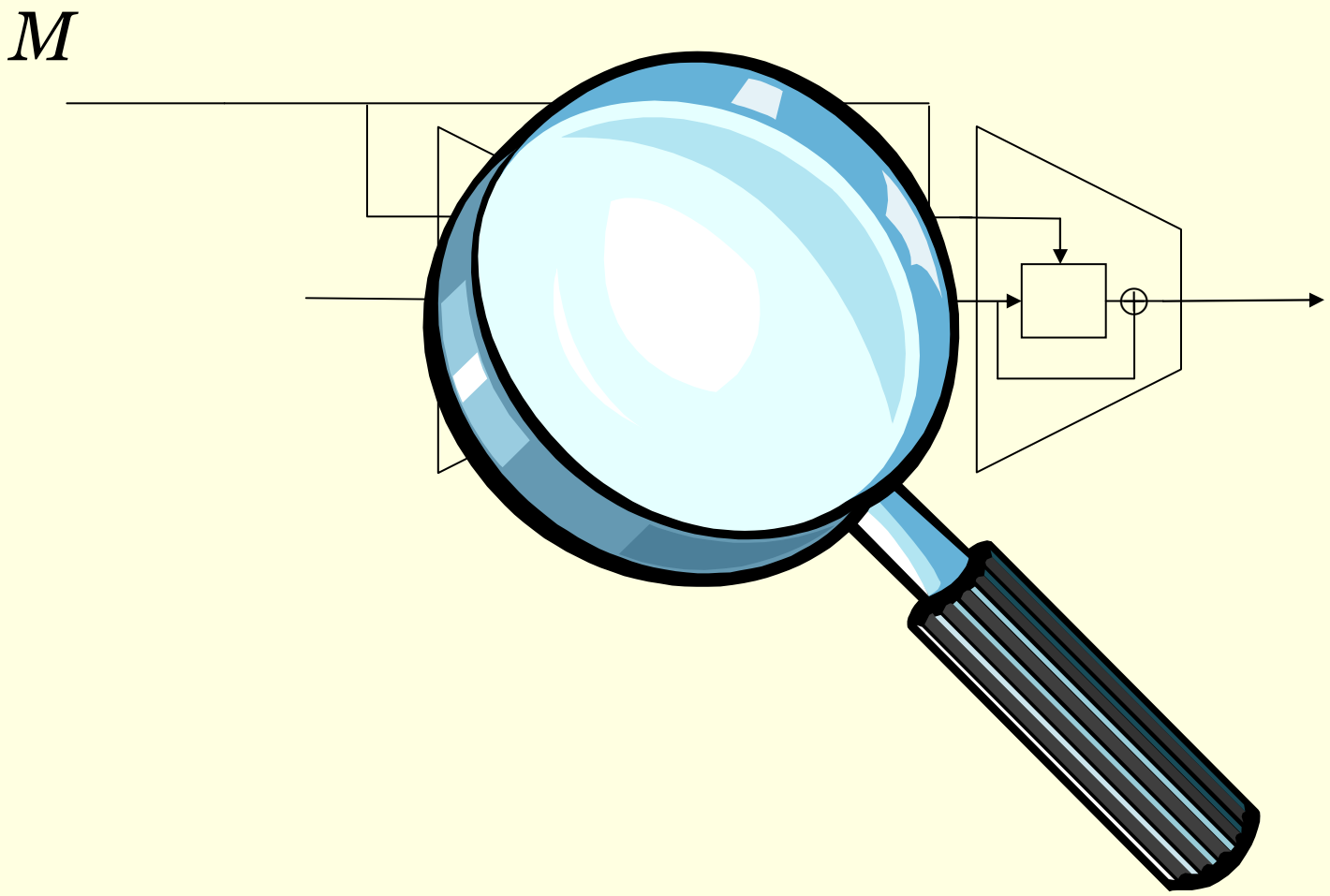
Overview of construction



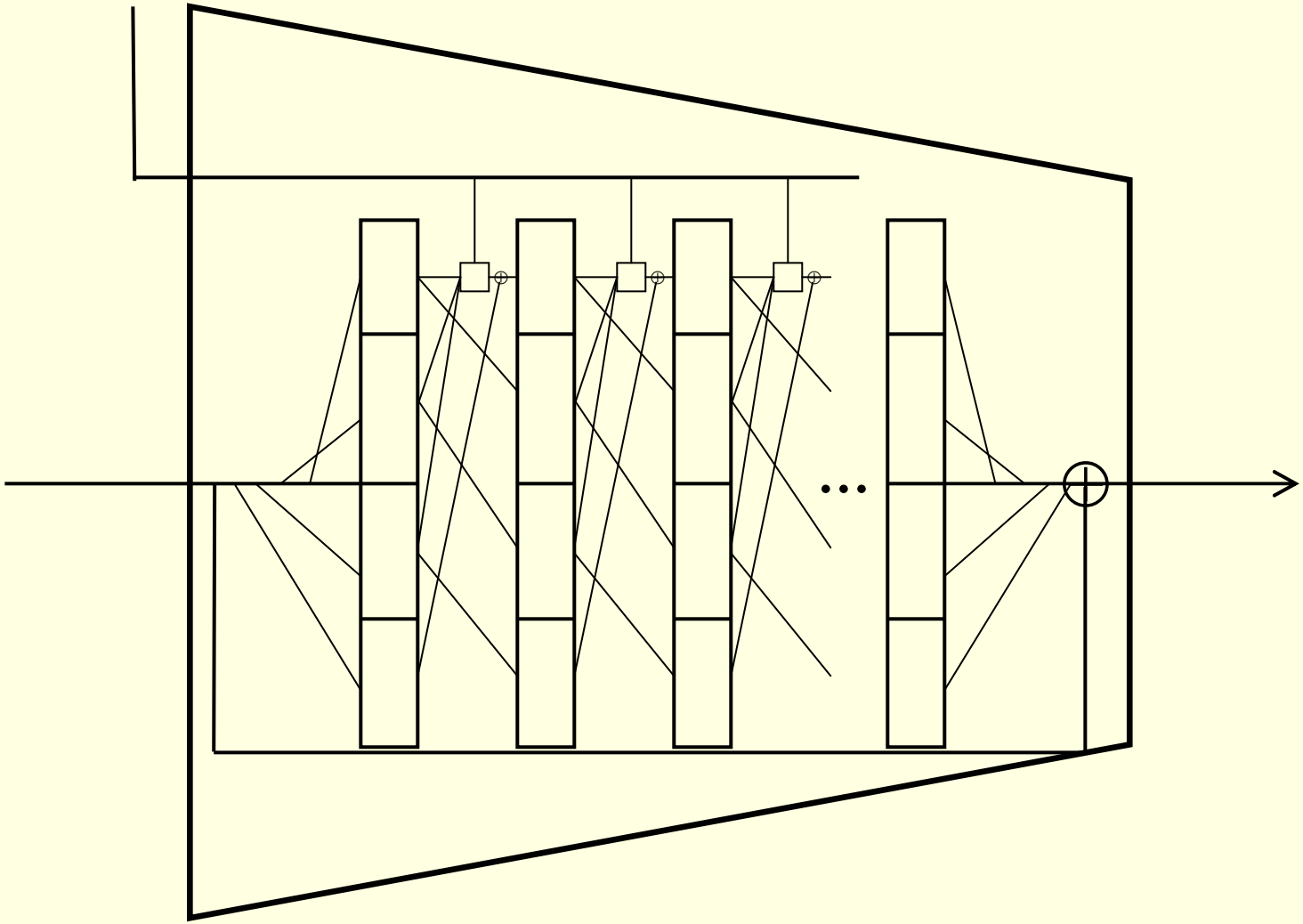
Overview of construction



Overview of construction



Overview of construction



Black box no more

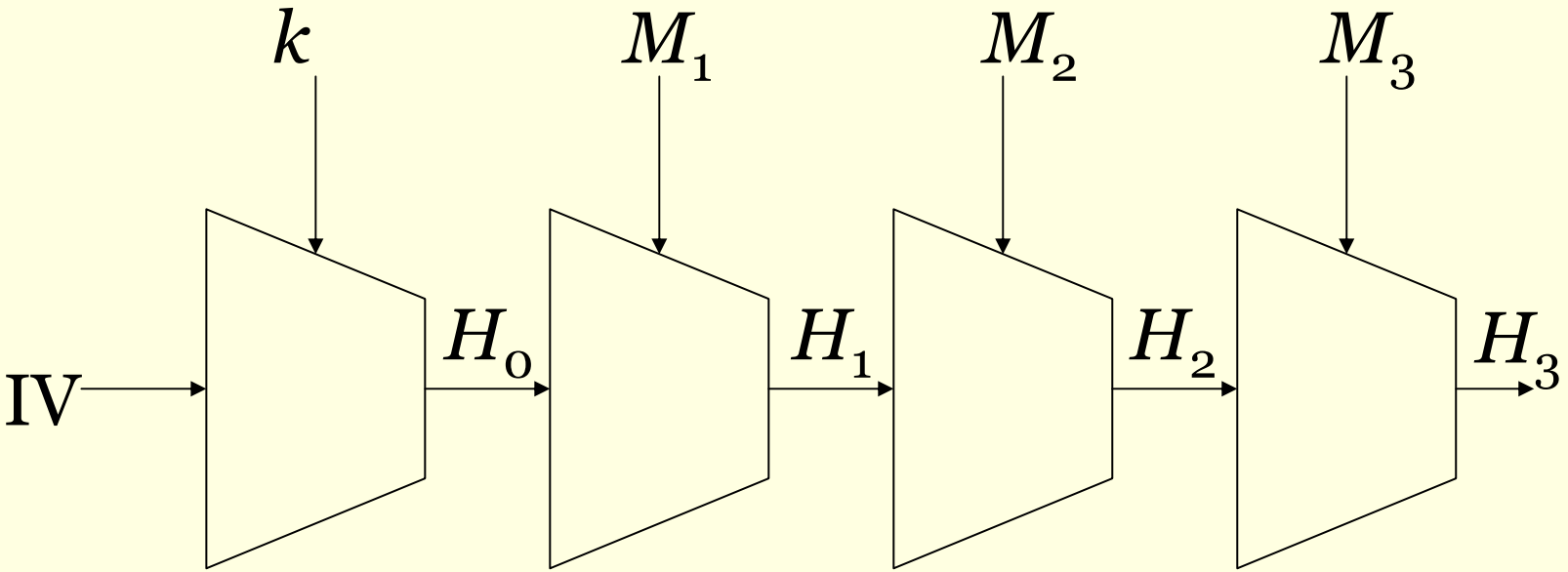
Outline

Hash functions:

- Definitions
- Constructions
- Attacks
- What to do

Authentication

MAC: $H_k(M) = H(k \,||\, M)$



$$H_k(M \,||\, M_3) = C(H_k(M), M_3)$$

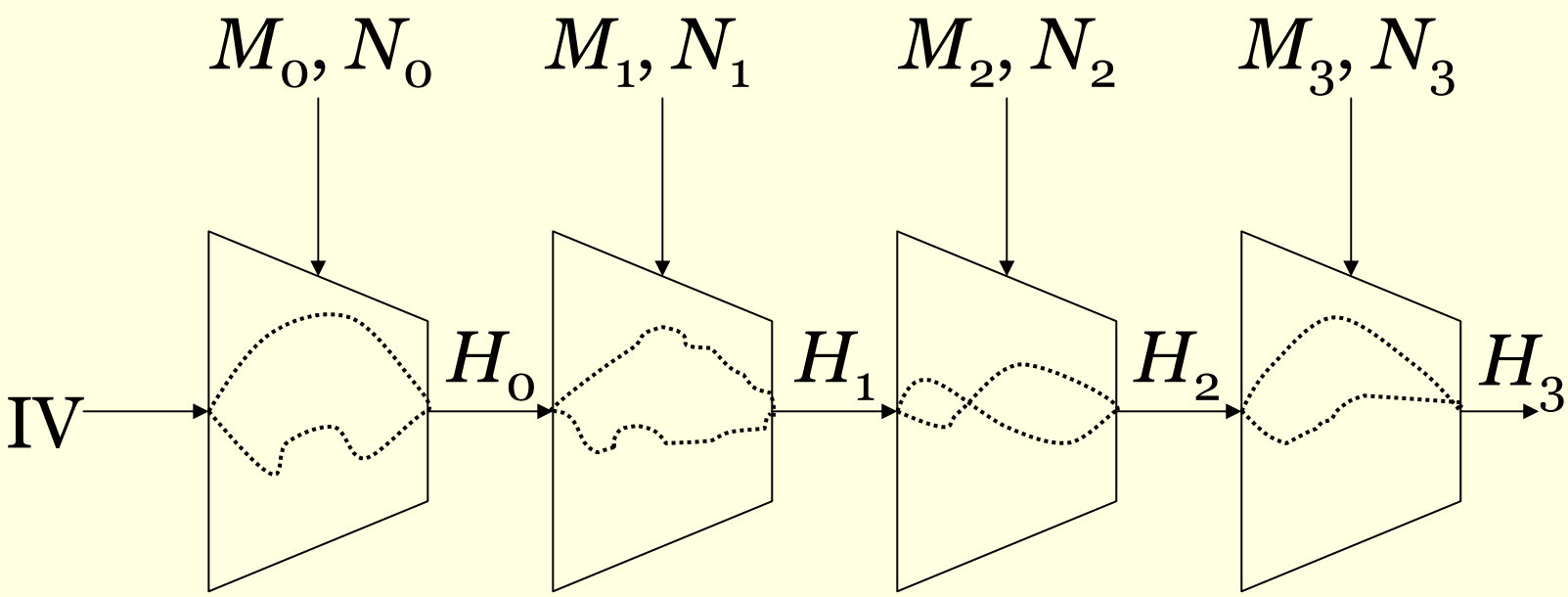
Cascading hash-functions

$F(M) = G_1(M) \,||\, G_2(M)$ – cascade

«If G_1 and G_2 are independent, then finding a collision for F requires finding a collision for both simultaneously (i.e., on the same input), which one could hope would require the **product of the efforts** to attack them individually.»

Handbook of Applied Cryptography

Joux attack

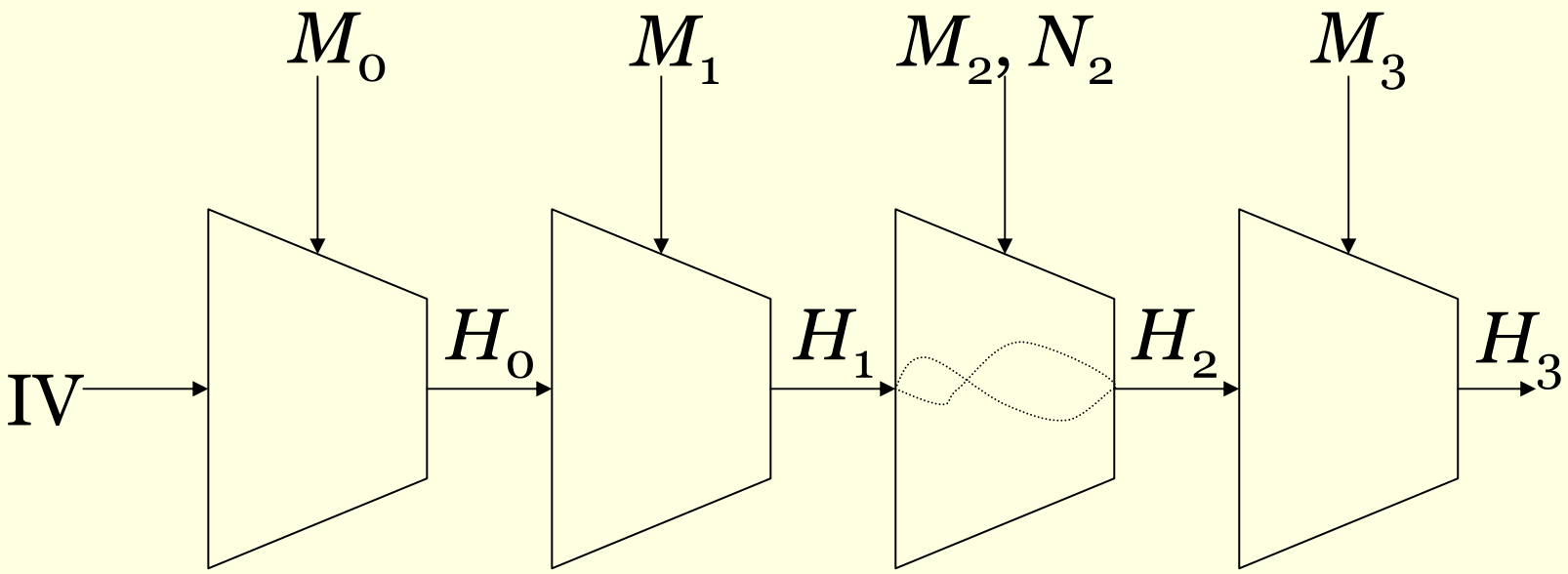


$k\ 2^{n/2}$ effort to find 2^k collisions

Joux attack

1. With $k \cdot 2^{n/2}$ effort, find 2^k collisions for G_1
2. If k = $n/2$, we may find **among them** collision M, N for G_2
3. $G_1(M) = G_1(N)$ and $G_2(M) = G_2(N)$

«Poisoned block»



Attacking Postscript

```
if (R2 == R1) then print("Dr. Jekyll")
              else print("Mr. Hyde")
```

Lenstra et al.: colliding X.509

X.509 certificate:

- version,
- serial number,
- signature algorithm identifier
- …
- RSA modulo
- …
- signature

Lenstra et al.: colliding X.509

X.509 certificate:

version,

serial number,

signature algorithm identifier

…

RSA modulo

…

signature

N_1 = <collision 1> 24519253

N_2 = <collision 2>24519253

Practice (second millennium)

name	**author**	**length**	**# rounds**	**year**
MD4	Ron Rivest	128 bit	48 rounds	1990
MD5	Ron Rivest	128 bit	64 rounds	1992
SHA0	NSA	160 bit	80 rounds	1993
SHA1	NSA	160 bit	80 rounds	1995

Practice (third millennium)

name	**author**	**word** (bits)	**length** (bits)	**rounds**	**year**
SHA-256	NSA	32	256	64	2002
SHA-384	NSA	64	384	80	2002
SHA-512	NSA	64	512	80	2002
Whirlpool	Barreto, Rijmen	—	512	10	2003

Speed

In CPU cycles/byte on PIII:

MD5	3.7
SHA1	8.3
SHA-256	21
SHA-512	40
Whirlpool	36

Attacks on MD4-MD5

hash	author	type	complexity	year
MD4	Dobbertin	collision	2^{22}	1996
	Wang et al.	collision	2^{8}	2005
MD5	dan Boer & Bosselaers	pseudo-collision	2^{16}	1993
	Dobbertin	free-start (chosen IV)	2^{34}	1996
	Wang & Yu	collision	2^{39}	2005

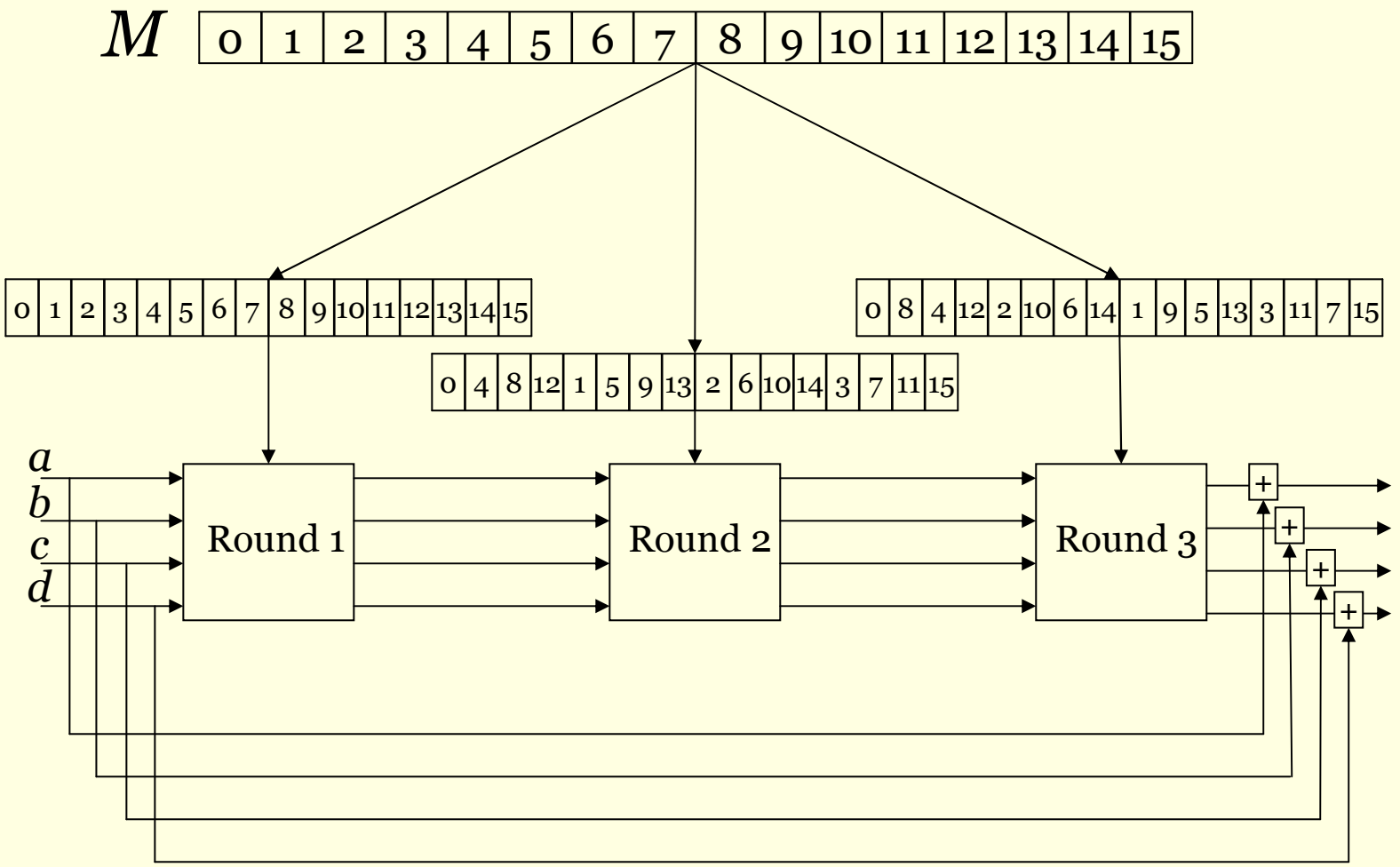
Attacks on MD4-MD5

hash	author	type	complexity	year
MD4	Dobbertin	collision	2^{22}	1996
	Wang et al.	collision	2^{8}	2005
MD5	dan Boer & Bosselaers	pseudo-collision	2^{16}	1993
	Dobbertin	free-start (chosen IV)	2^{34}	1996
	Wang & Yu	collision	2^{39}	2005

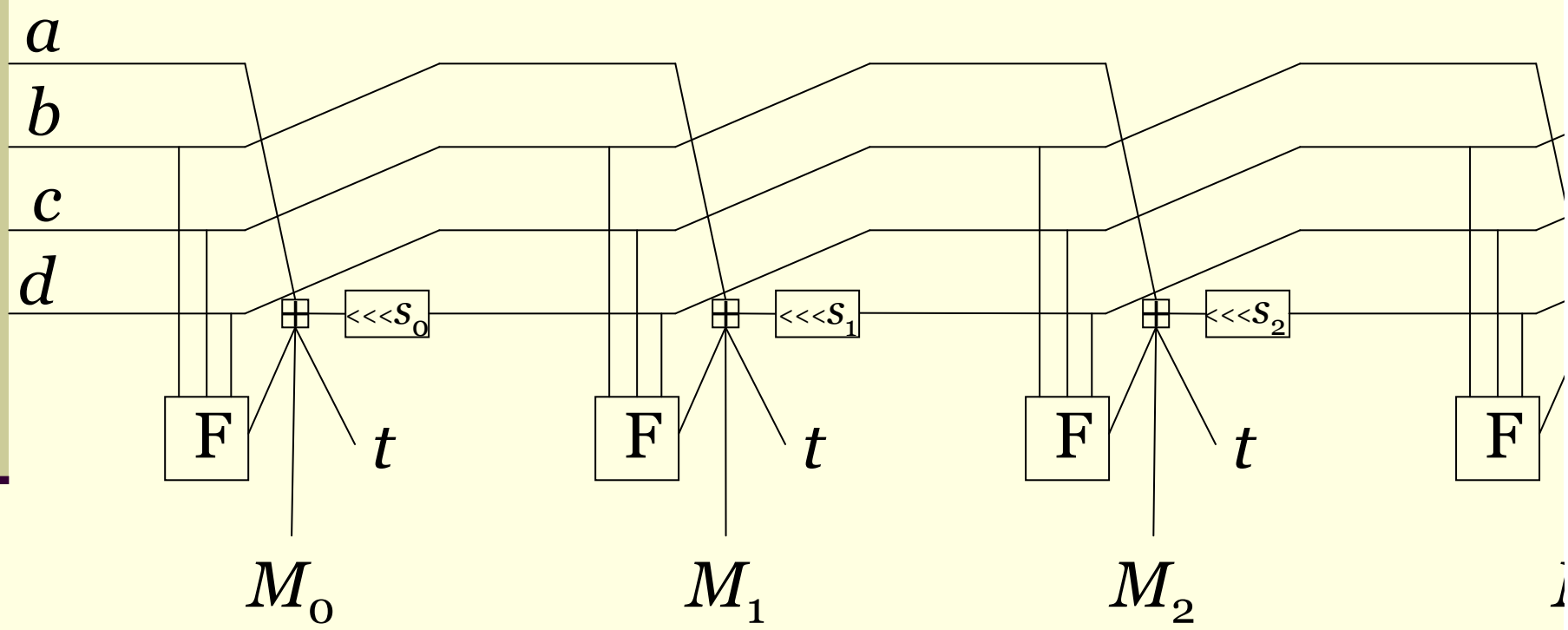
Attack on SHA-0,1

hash	author	type	complexity	year
SHA0	Chabaud & Joux	collision	2^{61} (theory)	1998
	Biham & Chen	near-collision	2^{56}	2004
	Biham et al.	collision	2^{51}	2005
	Wang et al.	collision	2^{39}	2005
SHA1	Biham et al.	collision (40 rounds)		2005
	Wang et al.	collision (58 rounds)	2^{33}	2005
	Wang et al.	collision	2^{63} (theory)	2005

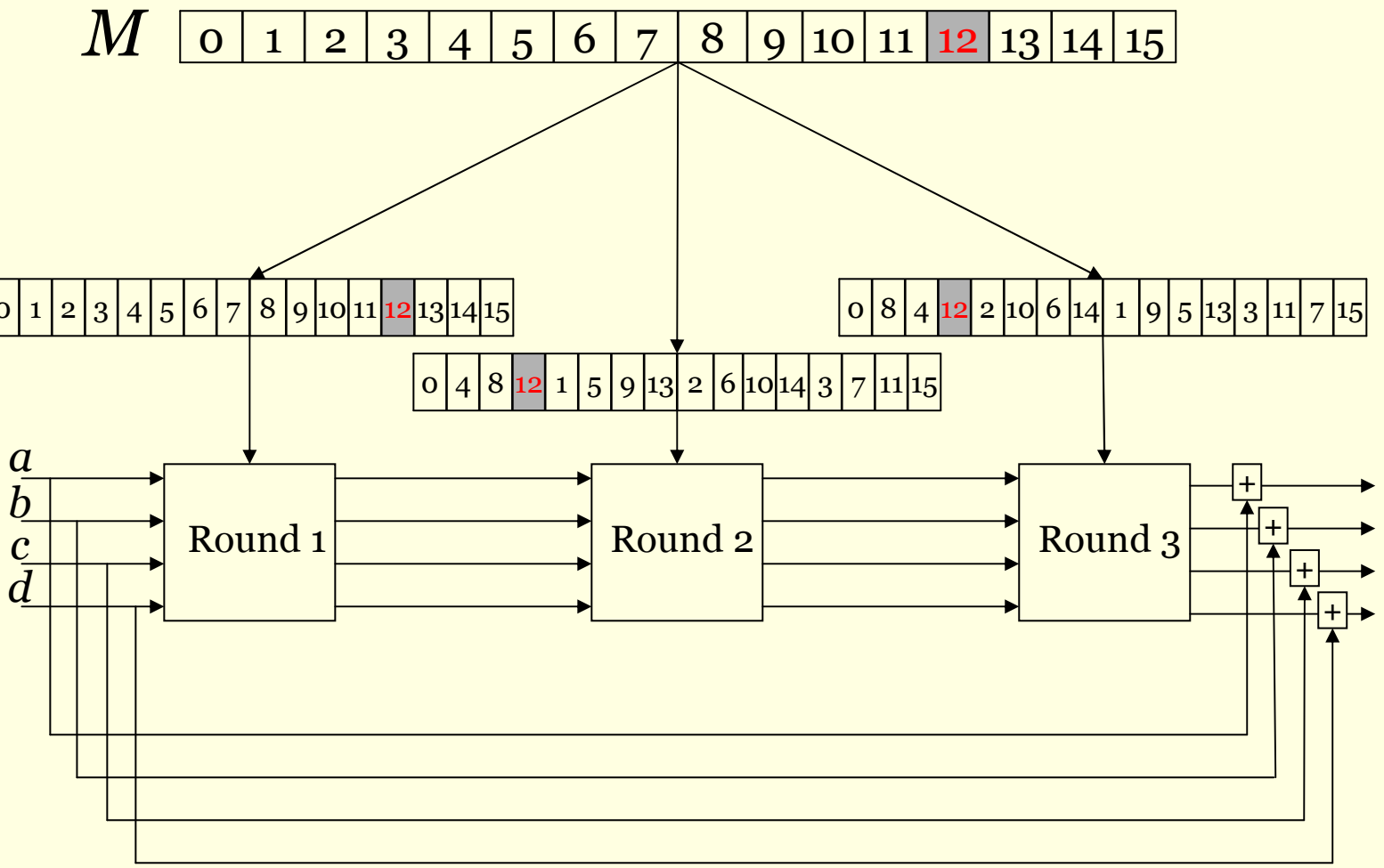
Detailed structure of MD4



Detailed structure of MD4



Introducing differentials



Outline

Hash functions:

- Definitions
- Constructions
- Attacks
- What to do

What is safe?

- Entropy extraction
- Randomization
- Cryptography
  - Signatures
  - Authentication

Entropy extraction

- Scenario I:

$H(g^{ab})$

- Scenario II:

IP, time, name, ... → GUID

Randomization

Pair-wise independent functions (Wegman & Carter):

$$f(x) = ax+b \bmod N$$

probability $f(x_1)=f(x_2)$ equals 1/N

$$f(x) = a_k x^k + a_{k-1} x^{k-1} + \ldots + a_1 x + a_0 \bmod N$$

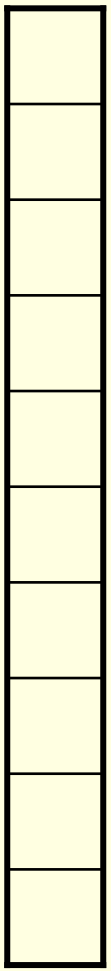
Randomization

- Scenario: hash-table

3 12 78 40 67 15

a, b, c - random

$H(x) = ax^2+bx+c \bmod N$



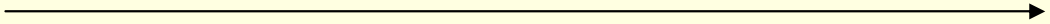
Cryptography

Authentication

Alice Bob

shared key K

M



$H_K(M)$

Eve: $M_1, H_K(M_1), M_2, H_K(M_2) \ldots \rightarrow M, H_K(M)$

Cryptography

Authentication

UMAC

John Black, Shai Halevi, Hugo Krawczyk, Ted Krovetz, and Phillip Rogaway, 1999

RFC: under way

for long messages ~1.5 cycle/byte

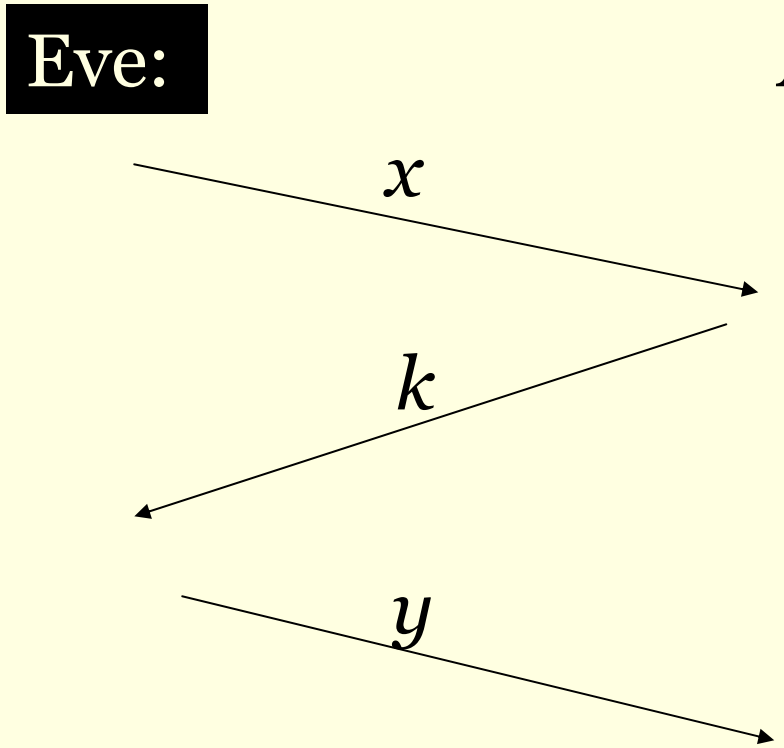
Cryptography

- Signatures
  theory: target-collision resistance
  practice: SHA1, SHA256, SHA512, Whirlpool

- Protocols, constructions
  ??????

Target-collision resistance

$H_k(M)$



$H_k(x)=H_k(y)$

Signing using TCR hashes

$H_k(M)$ – TCR

To sing M:

1. Generate k
2. Sign $k \,||\, H_k(M)$

Attack?

Find N so that

$$k \,||\, H_k(M) = k \,||\, H_k(N)$$

Questions?